

# Моделирование окраски органических соединений\*

Козеева О.О., Чухраев И.В., Дерюгина Е.О.

Московский государственный технический университет им. Н.Э. Баумана  
(Калужский филиал)

г. Калуга, Российская Федерация

[blueelectricat@gmail.com](mailto:blueelectricat@gmail.com), [igor.chukhraev@mail.ru](mailto:igor.chukhraev@mail.ru), [syvorova\\_eo@mail.ru](mailto:syvorova_eo@mail.ru)

**Аннотация.** Органические красители и пигменты широко применяются в промышленности и исследование их окраски с целью получения требуемых свойств может осуществляться посредством моделирования структуры химических соединений с помощью средств ЭВМ, что является задачей автоматизации соответствующего эксперимента. Для ее решения разработана информационная система, позволяющая осуществлять интерактивное конструирование структур соединений, оперативное и динамическое отображение результата в доступной визуальной форме и предоставляющая доступ к внешним базам химических данных для получения справочной информации. Таким образом, данный программный продукт решает недостатки существующих вычислительных средств. В статье рассмотрен принцип построения системы в целом, разработан алгоритм реализации расчетного модуля на примере определенного класса соединений, обозначены некоторые перспективы дальнейшего совершенствования программы.

**Ключевые слова:** компьютерное моделирование, органические соединения, органические красители, Python, RDKit.

## ВВЕДЕНИЕ

Органические красители и пигменты широко используются в различных отраслях промышленности. Пигменты входят в состав лакокрасочных материалов и придают им определённые свойства: собственно, оптические, изоляционные, стойкость к температурам, коррозии и воздействию кислот [1]. Пигменты имеют интенсивную окраску, а их органическое происхождение в некоторых случаях является аспектом их большей экологичности и безопасности по сравнению с неорганическими красителями. Таким образом, поиск структуры пигмента с нужным цветовым решением является немаловажным вопросом для ряда областей: от пищевой до машиностроения и медицины [2].

Растворы органических красителей применяются в качестве активной среды лазеров с узконаправленным монохроматическим излучением для проведения высокоточных операций, в том числе, для обработки материалов, а также при изготовлении микрочипов. Такие лазеры имеют ряд преимуществ по сравнению с газовыми и твердотельными, во многом благодаря сильным люминесцентным свойствам [3]. Подбор красителей позволяет получить излучение необходимой длины волны в широком диапазоне частот, так, одной из задач является разработка лазеров с перестраиваемой длиной волны излучения [4]. Также моделирование структуры красителей с определенными

ми спектральными характеристиками необходимо для разрешения проблем их химической неустойчивости и взаимодействия с растворами [5].

Также полупроводниковые свойства органических красителей обуславливают их использование в производстве гибких и тонких пленочных фотопанелей [6].

Программа, рассматриваемая в данной статье, позволяет выполнять первичный поиск решений перечисленных задач. Существующие программные продукты [7-9], предназначенные для проведения научных расчетов в различных областях исследований, потенциально имеют возможности для моделирования структуры органических соединений и определения их окраски, однако, зачастую, они основаны на квантово-химических расчетах либо на методах машинного обучения, что ведет за собой повышенные требования к вычислительным ресурсам ЭВМ. Поставленная задача не является целевой для данных систем, и они обладают избыточным функционалом, имеют сложности в эксплуатации с точки зрения гибкости системы ввода данных, так как даже незначительная ошибка при ручном вводе данных может существенно сказаться на результате. Корректная интерпретация результатов, их динамическое отображение, визуальная составляющая при конструировании молекул также являются одними из ключевых требований для системы, позволяющей решать поставленную задачу при обеспечении доступности эксплуатации и быстрого действия. На данный момент обозначенные требования не удовлетворяются существующими системами в полной мере.

Исследование окраски органических соединений является актуальной задачей синтеза пигментов и красителей с целью получения требуемых характеристик. Для этого применяются различные методы химического анализа и синтеза. В данной статье рассматривается разработка программы прогнозирования окраски соединений на основе положений структурной теории цветности. Согласно им, цвет соединений определяется наличием в составе вещества определенных функциональных подгрупп – хромофоров. Моделирование структуры соединений может осуществляться присоединением к исходной молекуле данных групп, удлинением цепи сопряженных связей, что влияет на распределение электронного заряда и, как следствие, эмиссионные свойства [10-12].

Целью разработки рассматриваемой программы является увеличение быстрого действия и улучшение наглядно-

\* Статья публикуется по рекомендации программного комитета Международной научно-практической конференции "Материаловедение и металлургические технологии" (RusMetalCon-2019), <https://rusmetalcon.susu.ru>

сти моделирования зависимости типа “структурасвойство”, обеспечение минимального времени отклика между поступлением входных данных и получением результата для решения задач поиска структур с нужными свойствами.

#### ОПИСАНИЕ СТРУКТУРЫ СИСТЕМЫ

В соответствии с приведенными выше требованиями к функциональным возможностям системы определяется ее структура, представленная на рис. 1.

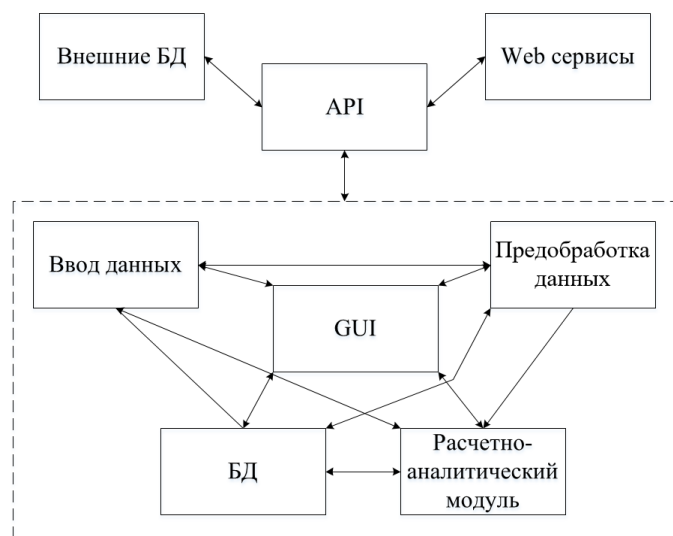


Рис. 1. Функциональная структура системы

Функции указанных модулей системы:

- 1) Предобработка данных: приведение данных вводимых в систему к необходимым форматам;
- 2) Расчетно-аналитический модуль: прогнозирование на основе расчетов по эмпирическим правилам;
- 3) GUI: визуальный ввод и вывод данных;
- 4) Базы данных (БД): хранение локальных данных;
- 5) Ввод данных: открытие, создание, сохранение файлов соответствующих форматов представления химических данных;
- 6) API: отправка и получение запросов к внешним ресурсам;
- 7) Внешние БД: получение необходимых справочных данных;
- 8) Web сервисы: предоставление услуг внешних сервисов.

Система разработана средствами языка Python, для которого разработано множество библиотек, необходимых для реализации программ различного предназначения, в том числе для проведения научных исследований в области химии.

Основной расчетный модуль реализован с использованием библиотеки RDKit, имеющей широкий набор методов для обработки химических данных и расчета различных характеристик. Более подробно разработка данного модуля будет рассмотрена ниже.

Графический интерфейс пользователя (GUI) предназначен для взаимодействия пользователя с системой, визуального ввода и вывода данных. Он разработан с помощью библиотеки PySide2. Выполняемые манипуляции по построению структур интерпретируются с помощью методов RDKit.

Модуль ввода данных представляет собой работу с файлами форматов .sdf, .mol, xls/xlsx, csv, а именно выполнение тривиальных задач по сохранению, созданию и открытию.

Данные программы, в том числе сохраненные результаты проведенных экспериментов, служебная информация, хранятся в локальной базе данных. Используемая СУБД – sqlite3.

Некоторые функции предоставляются внешними службами через API, реализованный на базе фреймворка Bottle. Так, справочная информация может быть получена подключением к базам данных PubChem посредством интерфейса PUG REST. Конвертация химических идентификаторов, например, SMILES в InChI, осуществляется отправкой запросов к ChemSpider API.

Таким образом, для реализации программы были выбраны средства с минимально необходимым функционалом, не составляющие дополнительную нагрузку при установке и эксплуатации системы и обеспечивающие требуемое быстродействие.

#### РАЗРАБОТКА РАСЧЕТНОГО МОДУЛЯ

В качестве расчетного метода вычисления спектральных характеристик органических красителей при реализации соответствующего модуля используются эмпирические правила Вудворда-Физера, которые позволяют рассчитать длину волны максимального поглощения для молекулы путем прибавления инкрементов к базовому значению, соответствующему определенному хромофору [13, 14]. Эти данные известны для различных классов соединений (диены, полиены, еноны и другие) и сведены в специальные таблицы [15]. В данной статье будет представлена разработка алгоритма для расчета длины волны максимального поглощения для полиенов. Для остальных классов соединений программа реализуется схожим образом.

Для вычисления длины волны максимального поглощения используется следующая формула:

$$\lambda_{\max} = 114 + 5 \cdot a + b \cdot (48 - 1,7 \cdot b) - 16,5 \cdot n - 10 \cdot x,$$

где  $\lambda_{\max}$  – длина волны, при которой наблюдается максимальное поглощение света;  $a$  – число алкильных заместителей;  $b$  – число сопряженных связей;  $n$  – число колец с внутрикольцевой мостиковой связью в сопряженной системе;  $x$  – число колец с внешней мостиковой связью в сопряженной системе [16]. В данном случае базовое значение хромофора равно 114.

На рис. 2 представлен разработанный алгоритм, функции `if_DoubleConj(smiles)` расчета числа сопряженных связей в молекуле для полиенов и реализованный метод-ами библиотеки RDKit.

На вход функции подается SMILES идентификатор сконструированной молекулы, информация из которого считывается методом `MolFromSmiles()`. Функция рассчитывает количество связей в молекуле методом `GetNumBonds()`, затем каждая связь берется по индексу (`GetBondWithIdx()`), методом `GetBondTypeAsDouble()` определяется ее тип, метод `GetIsConjugated()` возвращает TRUE, если связь является сопряженной, иначе FALSE, и если связь является двойной сопряженной, то счетчик инкрементируется. По окончании работы функция возвращает значение счетчика.

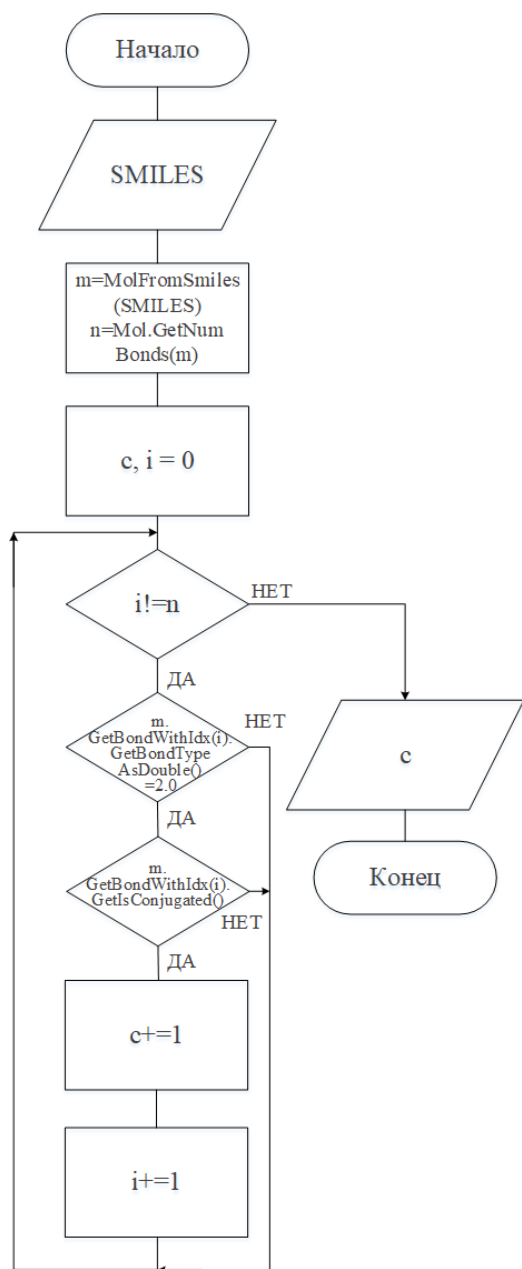


Рис. 2. Алгоритм функции if\_DoubleConj(smiles)

На рис. 3 представлен алгоритм функции TypeofBondInRing(smiles) для вычисления числа колец с внешней и с внутренней связью. Если оба атома углерода в двойной связи являются частью кольца, то такая связь называется внутрикольцевой, иначе – внешней.

В данной функции также на вход подается SMILES идентификатор, определяется количество связей. Затем проверяется тип каждой связи. Для двойных связей определяется, оба ли атома углерода являются частью кольца методом IsInRing(), который в таком случае возвращает TRUE. Если нет, то проверяется является ли хотя бы один из атомов частью кольца методом IsInRing(). Для этого методом GetBeginAtom() берется первый атом связи, а методом GetEndAtom() – последний. Затем осуществляется описанная проверка. Для каждого из случаев (внутрен-

ней и внешней связи) вводится счетчик, который инкрементируется в случае соответствующей идентификации. По завершении работы функция возвращает полученные значения счетчиков.

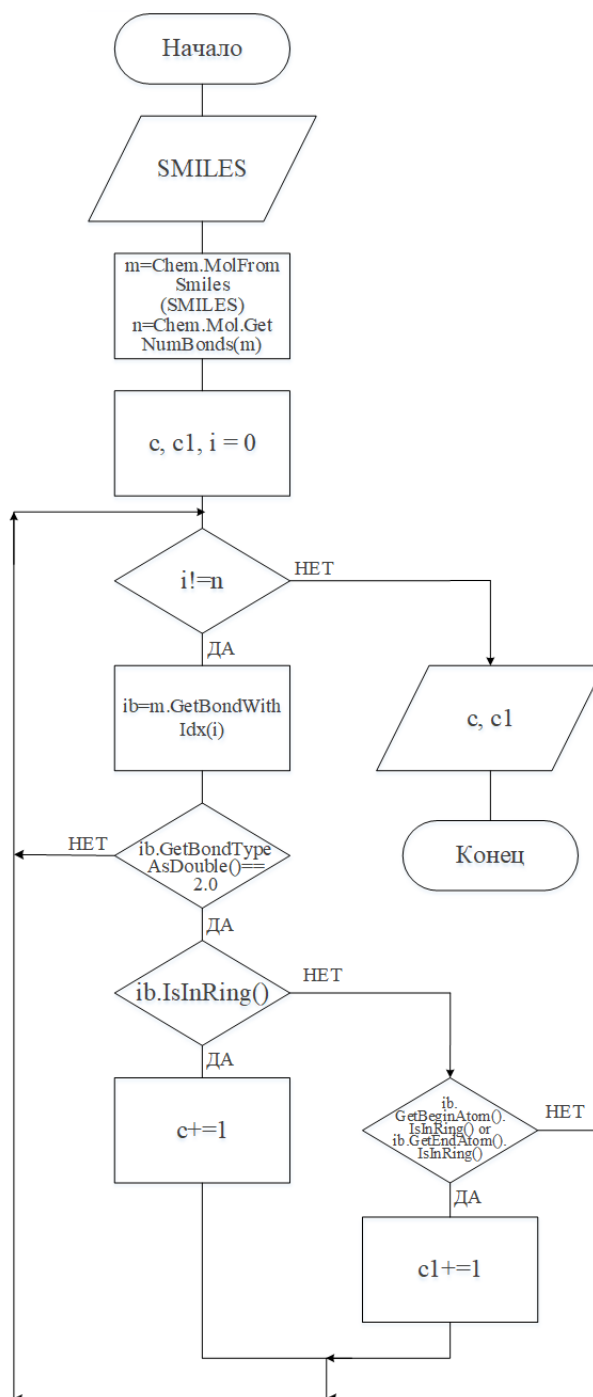


Рис. 3. Алгоритм функции TypeofBondInRing(smiles)

Для определения числа алкильных заместителей используется функция GetSubstruct разработанного ранее модуля поиска подструктур [17].

Таким образом, в общем виде расчетная функция abs\_max(smiles) для полиенов представляет собой следующий алгоритм операций (рис. 4).

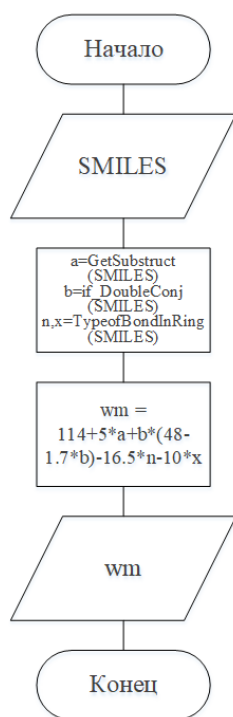


Рис. 4. Алгоритм функции abs\_max(smiles)

На вход функции подается идентификатор сконструированного соединения SMILES, который затем передается на обработку функциям, описанным выше. Полученные параметры подставляются в формулу для вычисления длины волны максимального поглощения, которая возвращается функцией в качестве результата. Для качественного определения цвета необходимо сопоставить полученной длине волны максимального поглощения дополнительный цвет [18] в соответствии с табл.

Таблица

Дополнительные цвета

Поглощаемая длина волны, нм	Цвет	Дополнительный цвет
400-450	фиолетовый	желто-зеленый
450-500	синий	желтый
500-550	зеленый	красный
550-600	желтый	синий
600-650	оранжевый	голубой
650-700	красный	зелено-голубой

Таким образом, с помощью методов RDKit была реализована программа расчета спектральных характеристик органических красителей.

Для определения точности расчетов анализировалось количество верных прогнозов: если результат находится в диапазоне  $\Delta L = \pm 50$  от истинного значения, то можно считать его допустимым. Границы ошибки были определены следующим образом: видимый спектр находится в диапазоне 400-700 нм; для классификации выбрано разбиение на 6 групп: фиолетовый, синий, зеленый, желтый, оранжевый, красный, и, таким образом, границы отнесения к определенному цвету составляют 50 нм. Метод определе-

ния диапазона допустимой погрешности результата может быть значительно улучшен, например, устранением жесткой линейности классификации.

#### ЭКСПЛУАТАЦИЯ СИСТЕМЫ

Реализованная программа имеет графический интерфейс (рис. 5), который позволяет осуществлять конструирование структур химических соединений.

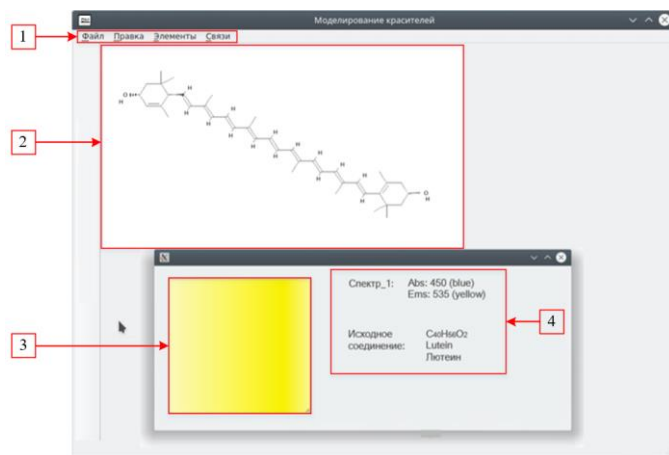


Рис. 5. Интерфейс программы

Рабочее окно содержит следующие элементы:

- 1) Главное меню;
- 2) Канва;
- 3) Градиентное отображение результата;
- 4) Выходные характеристики.

Главное меню имеет стандартный пункт “Файл” с опциями открытия, сохранения файлов в форматах .sdf, .mol, “Правка” для отмены и повтора действий, “Элементы” для выбора химических элементов, “Связи” для выбора связей различного типа.

В качестве основного метода расчета используется рассмотренная функция abs\_max. Градиентное изображение результата формируется с учетом погрешности, определенной для каждого конкретного класса соединения.

После выполнения действий по конструированию молекулы в дочернем окне вывода результата отображаются выполненные изменения (рис. 6).



Рис. 6. Окно отображения результата

Проверка работоспособности программы при отслеживании выходных данных расчетных функций продемонстрирована на примере бета-каротина (рис. 7).



```
s='CC1=C(C(CCC1)C)C/C=C/C(=C/C=C/C(=C/C=C/C=C  
/C=C/C=C(/C=C/C2=CCCC2(C)C)\C)\C)/C/'  
if_DoubleConj(s)  
  
11  
TypeofBondInRing(s)  
  
(2, 0)  
  
abs_max(s)  
  
453.3
```

Рис. 7. Результаты работы функций для бета-каротина

Полученные значения соответствуют истинным [16].

#### ЗАКЛЮЧЕНИЕ

В данной работе была рассмотрена реализация программы прогнозирования окраски органических красителей и пигментов, описана ее общая структура и средства реализации. Алгоритм, заложенный в основе расчетного модуля программы, дает точный результат со средней погрешностью 5 нм. Программа может быть дополнительно усовершенствована, например, реализацией более гибкой интерактивной справочной системы, в которой при конструировании молекулы выводятся подсказки для пользователя, как те или иные изменения в структуре влияют на спектральные свойства вещества.

В целом же, система представляет собой законченное решение поставленной задачи и может применяться как полностью самостоятельный программный продукт, предназначенный для первичного поиска структур с нужными свойствами.

#### ЛИТЕРАТУРА

1. Herbst W. Industrial organic pigments: production, properties, applications. / W. Herbst, K. Hunger. – John Wiley & Sons, 2006.
2. Faulkner E.B. High performance pigments. / E.B. Faulkner, R.J. Schwartz. – John Wiley & Sons, 2009.
3. Иванов А.Ю. Эффективный режим лазерной генерации смеси органических красителей на длинах волн 610 и 670 nm // Письма в ЖТФ. – 1999. – Т. 25, №. 8.
4. Кардашев Г.А. Физические методы интенсификации процессов химической технологии. – М.: Химия, 1990. – 208 с.

5. Тарасов Л.В. Физика процессов в генераторах когерентного оптического излучения. – М.: Радио и связь, 1981. – 440 с.

6. Ashford D.L. Molecular chromophore–catalyst assemblies for solar fuel applications // Chemical reviews. – 2015. – Vol. 115, No. 23. – P. 13006-13049.

7. Software and resources for computational medicinal chemistry / Liao C. et al. // Future medicinal chemistry. – 2011. – Vol. 3, No. 8. – P. 1057-1085. DOI: 10.4155/fmc.11.63

8. Pirhadi S. Open source molecular modeling / S. Pirhadi, J. Sunseri, D.R. Koes // Journal of Molecular Graphics and Modelling. – 2016. – Vol. 69. – P. 127-143. DOI: 10.1016/j.jmgn.2016.07.008

9. Koike A. Comparison of methods for chemical-compound affinity prediction // SAR and QSAR in Environmental Research. – 2006. – Vol. 17, No. 5. – P. 497-514. DOI: 10.1080/10629360600934168

10. Теренин А.Н. Фотохимия красителей и родственных органических соединений. – М.: Рипол Классик, 2013. – 353 с.

11. March J. Advanced organic chemistry: reactions, mechanisms, and structure. – John Wiley & Sons, 1992. – 1820 p.

12. Шретер В. Химия. / В. Шретер, К.Х. Лаутеншлегер, Х. Бибрак. – М.: "Химия", 2000.

13. Kalsi P.S. Spectroscopy of organic compounds. – New Age International, 2007.

14. Silverstein R.M. Spectrometric identification of organic compounds. – John Wiley & Sons.

15. Pretsch E. Tables of spectral data for structure determination of organic compounds. – Springer Science & Business Media, 2013.

16. Sharma Y.R. Elementary organic spectroscopy. – S. Chand Publishing, 2007.

17. Козеева О.О. Разработка на языке Python модуля поиска подструктур в химических соединениях / О.О. Козеева, И.В. Чухраев, А.В. Родионов // Электромагнитные волны и электронные системы. – 2018. – Т. 23, № 3. – С. 57-61.

18. Dickerson R.E. Chemical principles. / R.E. Dickerson, H.B. Gray, G.P. Haight. – The Benjamin/Cummings Publishing Company, Inc., 1979.

DOI: 10.24892/RIJIE/20200109

## Colour Modeling of Organic Compounds

Kozeeva O.O., Chukhraev I.V., Deriugina E.O.

Kaluga Branch of Bauman MSTU

Kaluga, Russian Federation

[bluelectricat@gmail.com](mailto:bluelectricat@gmail.com), [igor.chukhraev@mail.ru](mailto:igor.chukhraev@mail.ru), [syvorova\\_eo@mail.ru](mailto:syvorova_eo@mail.ru)

**Abstract.** Organic dyes and pigments are widely used in industry and researches of their colour features can be carried out by modeling the structure using computer tools in order to obtain the required properties. To this end, the program that provides interactive design of compound structures, dynamic display of the result in an accessible visual form and access to external chemical databases for obtaining reference information has been developed. Thus, this software product solves the shortcomings of

existing computational tools. The article describes the methods of building the system as a whole, process of developing an algorithm for implementing the calculation module and an additional method used to calculate the spectral characteristics, outlines the prospects for further improvement of the program.

**Keywords:** computer modeling, organic compounds, organic dyes, Python, RDKit.

REFERENCES

1. Herbst W., Hunger K. Industrial organic pigments: production, properties, applications, John Wiley & Sons, 2006.
2. Faulkner E.B., Schwartz R.J. High performance pigments, John Wiley & Sons, 2009.
3. Ivanov A.Yu. Anomalous backscattering of high-power laser radiation by the flat surface of a solid, *Technical Physics Letters*, 1999, vol. 25, no. 8, pp. 645-649. DOI: 10.1134/1.1262584
4. Kardashev G.A. *Fizicheskie metody intensivatsii protsessov khimicheskoy tekhnologii* [Physical methods of intensification of chemical technology processes], Moscow, Chemistry, 1990, 208 p. (in Russ.)
5. Tarasov L.V. *Fizika protsessov v generatorakh kogerentnogo opticheskogo izlucheniya* [Physics of processes in generators of coherent optical radiation], Moscow, Radio and communications, 1981, 440 p. (in Russ.)
6. Ashford D.L. Molecular chromophore–catalyst assemblies for solar fuel applications, *Chemical reviews*, 2015, vol. 115, no. 23, pp. 13006-13049.
7. Liao C. et al. Software and resources for computational medicinal chemistry, *Future medicinal chemistry*, 2011, vol. 3, no. 8, pp. 1057-1085. DOI: 10.4155/fmc.11.63
8. Pirhadi S., Sunseri J., Koes D.R. Open source molecular modeling, *Journal of Molecular Graphics and Modelling*, 2016, vol. 69, pp. 127-143. DOI: 10.1016/j.jmgm.2016.07.008
9. Koike A. Comparison of methods for chemical-compound affinity prediction, *SAR and QSAR in Environmental Research*, 2006, vol. 17, no. 5, pp. 497-514. DOI: 10.1080/10629360600934168
10. Terenin A.N. *Fotokhimiya krasiteley i rodstvennykh organicheskikh soedineniy* [Photochemistry of dyes and related organic compounds], Moscow, Ripol Classic, 2013, 353 p. (in Russ.)
11. March J. *Advanced organic chemistry: reactions, mechanisms, and structure*, John Wiley & Sons, 1992, 1820 p.
12. Schröter W., Lautenschleger K.Kh., Bibrak H. *Khimiya* [Chemistry], Moscow, Chemistry, 2000. (in Russ.)
13. Kalsi P.S. *Spectroscopy of organic compounds*, New Age International, 2007.
14. Silverstein R.M. *Spectrometric identification of organic compounds*, John Wiley & Sons.
15. Pretsch E. *Tables of spectral data for structure determination of organic compounds*, Springer Science & Business Media, 2013.
16. Sharma Y.R. *Elementary organic spectroscopy*, S. Chand Publishing, 2007.
17. Kozeeva O.O., Chukhraev I.V., Rodionov A.V. Development of a substructure search module in Python [Razrabotka na yazyke Python modulya poiska podstruktur v khimicheskikh soedineniyakh], *Elektromagnitnye volny i elektronnye sistemy* [Electromagnetic waves and electronic systems], 2018, vol. 23, no. 3, pp. 57-61. (in Russ.)
18. Dickerson R.E., Gray H.B., Haight G.P. *Chemical principles*, The Benjamin/Cummings Publishing Company, Inc., 1979.

**Библиографическое описание статьи**

Козеева О.О. Моделирование окраски органических соединений / О.О. Козеева, И.В. Чухраев, Е.О. Дерюгина // *Машиностроение: сетевой электронный научный журнал*. – 2020. – Т.8, №1. – С. 50-55. DOI: 10.24892/RIJE/20200109

**Reference to article**

Kozeeva O.O., Chukhraev I.V., Deriugina E.O. Colour modeling of organic compounds, *Russian Internet Journal of Industrial Engineering*, 2020, vol.8, no.1, pp. 50-55. DOI: 10.24892/RIJE/20200109